2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)
April 4-7, 2018, Washington, D.C., USA

A TIME DOMAIN CLASSIFICATION OF STEADY-STATE VISUAL EVOKED POTENTIALS USING DEEP RECURRENT-CONVOLUTIONAL NEURAL NETWORKS

Mohamed Attia[⋆] *Imali Hettiarachchi* [⋆] *Mohammed Hossny*[⋆] *Saeid Nahavandi* [⋆]

[⋆] Institute For Innovation and Research, Deakin University, Australia

ABSTRACT

Steady-State Visual Evoked Potential (SSVEP) is one of the popular methods of brain-computer interfacing (BCI). It is used to translate the Electroencephalogram (EEG) signals into actions or choices. The main challenge in processing the SSVEP signal recognition is finding an appropriate intermediate representation to facilitate the classification task afterwards. In the literature, frequency domain analysis was extensively adopted as an intermediate representation for SSVEP classification. In this presented paper, we propose a deep learning model that uses a hybrid architecture based on Convolutional and Recurrent Neural Networks to classify SSVEP signals in the time domain directly. We achieved accuracy 93.59% compared to 87.40% for the state-of-the-art method: canonical correlation analysis in the frequency domain. The proposed architecture facilitates the real-time classification of SSVEP signals in the time domain for real-time applications such as robot cars and exoskeletons.

Index Terms— EEG, BCI, SSVEP, CNN, RNN, Time Domain.

1. INTRODUCTION

Brain computer interfacing (BCI) is one of the most active interdisciplinary research domains. It incorporates knowledge from different domain such as neuro-science, rehabilitation medicine, and machine learning [1]. BCI research aims to translate the brain signals into actions in order to restore some of the lost abilities for people with severe neuro-muscular disabilities, such as ALS patients [2].

Steady-State Visual Evoked Potential (SSVEP) is an electro-physiological signal of the brain, developed as periodic responses to the visual stimulation of the eye by a selected frequency. These responses are captured in the EEG signals with dominant frequency response at the selected frequency and its harmonics. SSVEP has many advantages such as applicability to the majority of the users with disabilities, high information transfer rate (ITR), needs small number of EEG electrodes and has reasonable signal to noise ratio [3, 4]. Many studies were based on the analysis of the EEG signal in the frequency domain to facilitate the detection of the main frequency response and its harmonics.

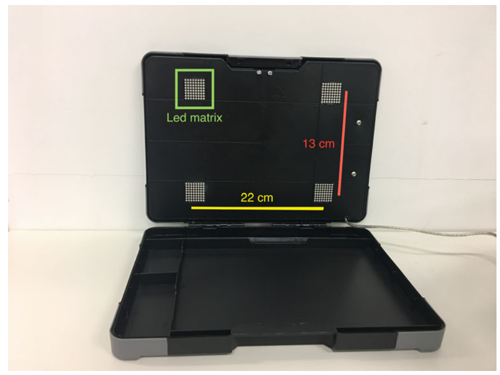

Fig. 1: The led matrix configuration. Each matrix blinks at different frequency: 6Hz, 7Hz, 8Hz, and 9Hz. Horizontal spacing between led matrix is 22 cm and vertical spacing is 13 cm.

Deep learning has been introduced as a self-learning methodology to solve many classification and regression tasks including image, video and speech. Convolutional neural networks (CNNs) are one of the most popular architectures for classification of images and videos. It is capable of extracting hierarchical representations for the input images. These self-learned features have great abilities for generalization and robustness to spatial translations. Recurrent neural networks have been used extensively to capture temporal information such as speech recognition and video classification.

The main contribution is the classification of the SSVEP signals using hybrid deep recurrent-convolutional architecture in the time domain as opposed to the frequency domain in the state-of-the-art method, such as canonical correlation analysis [5].

2. RELATED WORK

For detection of the SSVEPs, many studies used canonical correlation analysis (CCA) to detect the dominant frequency. It is a statistical method using multivariate analysis to model the relationships between two or more variables. Thus, it can detect the dominant frequency and the harmonics present in the SSVEP signals. Vu *et al.* proposed a modified architecture for CCA based on Stacked Auto Encoder (SAE) to the

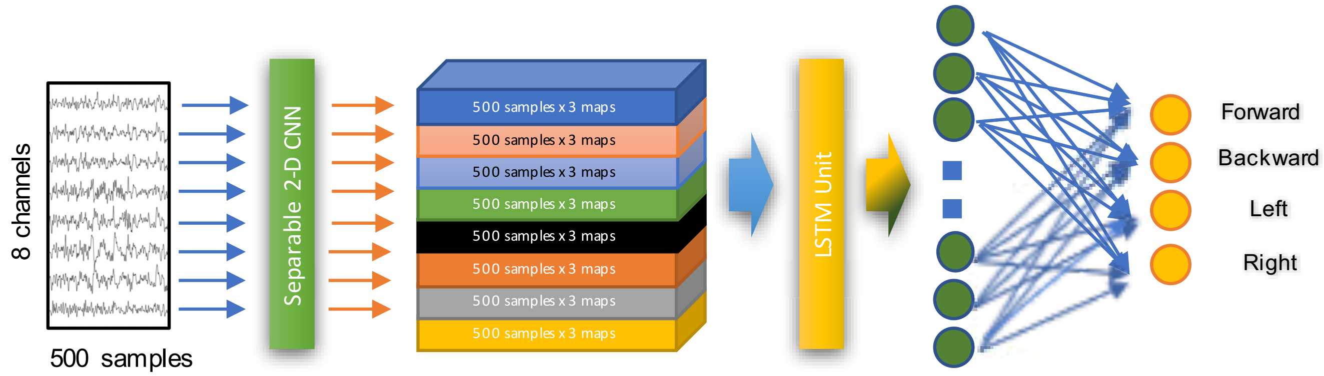

Fig. 2: Proposed network for SSVEP signal classification. It consists of depth-wise CNN layer, RNN layer and Fully connected with softmax activation.

maximize the correlation between the two sets of datasets [6]. SAEs are used to maximize the correlation through complex nonlinear transformations of the two sets [7].

Zhang *et al.* proposed a novel methodology based on multi-way extension of CCA [5]. In [8], they enhanced the accuracy of classification by using phase-constrained CCA. Zhang *et al.* proposed a novel method based on the sparsity constraint of the least absolute shrinkage and selection operator (LASSO) to extract more discriminating features from SSVEP signals [9]. In [10], Zhang *et al.* used multivariate synchronization index (MSI) as an index for decoding stimulus frequency by estimating the synchronization between two signals. In [11], they proposed a classifier using deep CNN to detect the dominant frequency using Fast Fourier transform of the EEG signal. In [12], they constructed simulated spectral topographical maps from the time series data by Fast Fourier transform. Then, they classified the maps using a CNN-based network.

Despite the great results obtained by the aforementioned methods, The input signal is buffered and converted into intermediate representation, such as the frequency domain or conversion into simulated topographical maps. Therefore, we propose a novel architecture for real-time classification of the SSVEP signals in the time domain using CNN and RNN.

3. PROPOSED METHOD

In the presented work, we propose a hybrid deep architecture that consists of a layer of depth-wise CNN and another one of RNN, as shown in Fig. 2. The Depth-wise CNN layer is responsible for extraction of the features from the input EEG signal. It performs a 2-D convolution with separable filters to capture features from each EEG channel independently from the others. Each convolutional kernel extract 3 feature maps from the input signal independently from the other seven channels. We chose the kernel size to be 3 with a stride of 1 sample. Then, the RNN layer is used to capture the temporal relationship between the extracted features. Schmidhuber *et al.* proposed Long Short-Term Memory (LSTM) as a recurrent neural network layer that capture the temporal relationships in sequences, through an internal memory and gated inputs/outputs [13]. In the proposed model, The feature maps extracted from the CNN layer are fed into the RNN layer with 128 LSTM units. Each unit computes a hidden vector sequence and an output vector, according to Eqn. (1- 6).

$$
\begin{aligned}
o_t &= \sigma(W_o * d_t + U_o * z_{t-1}) && (1)\\
z_t &= o_t * \tanh(C_t) && (2)\\
C_t &= f_t * C_{t-1} + A_t * \hat{C}_t && (3)\\
\hat{C}_t &= \tanh(W_c * p_t + U_c * z_{t-1}) && (4)\\
h_t &= \sigma(W_f * d_t + U_f * z_{t-1}) && (5)\\
r_t &= \sigma(W_A * d_t + U_A * z_{t-1}) && (6)
\end{aligned}
$$

where h_t is the forget gate, r_t is the reset gate, C_t is the memory cell and $\hat{C}_t$ is the exposure gate for the memory cell C_t. A softmax non-linearity layer is utilized to estimate the probability distribution over the classes. It is responsible for the conversion of arbitrary values of the output last layer $y \in \mathbb{R}^C$ to normalized probability prediction $p \in \mathbb{R}^C$ by exponentiation and normalization, as follows in Eqn. 7:

$$
p_i = \frac{e^{y_i}}{\sum_{c=1}^{C} e^{y_c}} \qquad (7)
$$

where $i, c \in \{1, 2, ..., C\}$ and C is the total number of classes. In our case, we have four classes C: forward, backward, right and left directions.

The proposed implementation is based on the architecture adopted in [14, 15]. However, We used only one layer of each of the CNN and RNN to avoid over-fitting of the model, due to the relatively limited number of training sample. It is also hard to augment the data by adding noise due to the nature of the signal which has low signal-to-noise ratio.

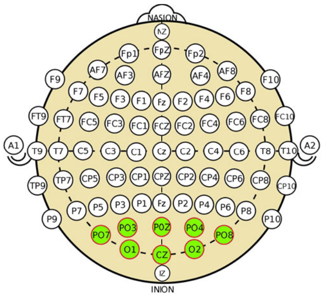

Fig. 3: The EEG electrode placement map. The selected channels were: 'PO7','PO3','POZ','PO4','PO8','01','OZ' and 'O2'.

4. EXPERIMENTS

The dataset was collected from four participants, including three male participants and one female participant. We were granted the ethics approval by the Human Ethics Advisory Group (HEAG) of the Faculty of Science Engineering and Built Environment, Deakin University, Australia. A full-written consent was obtained from each participant before the start of each experiment. The female participant wore eye glasses for corrected vision. The experiment was designed to control a robot car. We limited the car controls to four directions forward, backwards, turn left and turn right. We prepared a panel that consists of four LED matrices of the size 62(W) $\times$ 32(H) mm. Each matrix is designated to a direction, as shown in Fig. 1. In order to produce the visual stimulus for the SSVEP: the LED matrices are blinking at frequencies 6Hz, 7Hz, 8Hz and 9Hz. The led matrices are controlled by an Arduino Uno.

The data collection was done in two phases, a training phase and a test phase. The training phase consists of 60 trials in total. Each direction was presented to the participant 15 times per run in a randomized order. At the start of each trial of the training phase, each led matrix is fully light up for two seconds to attract the participant attention. Then, for the next seven seconds, the matrix start blinking at the designated frequency for that direction. Afterwards, the LED matrix is off for two seconds. In testing phase, all LEDs were blinking simultaneously.

We captured the EEG signal using 8 channels: 'PO7', 'PO3', 'POZ', 'PO4', 'PO8', 'O1', 'OZ' and 'O2', as shown in Fig. 3 using a wireless EEG amplifier, g.Nautilus (g.tec medical engineering GmbH, Austria). The data was recorded with an earlobe reference and a ground electrode at the location AFz, between FPz and Fz. The sampling frequency for the captured data was 250 Hz. The band-pass filter was set to 0.5 - 30 Hz and the notch filter was set to 50 Hz. The g.Nautilus is designed with active electrode technology to record high quality EEG data. During the recording, the electrode impedance was maintained below 30 kΩ for the male subjects. However for the female subject with longer hair, the impedance at several electrode sites were maintained at 50 kΩ level. During the training of the proposed model, we randomly split each dataset into train and test sets by the ratio 85% for training and 15% for testing.

Participant	Dataset	CCA	Proposed
Subject1	Train	87.51%	**95.38%**
	Test	86.88%	**96.26%**
Subject 2	Train	**91.65%**	89.64%
	Test	85.21%	**94.05%**
Subject 3	Train	97.9%	**98.31%**
	Test	96.03%	**96.46%**
Subject 4	Train	76.77%	**87.23%**
	Test	77.29%	**91.44%**
Average	-	87.40%	**93.59%**

Table 1: Results for the accuracy of the Proposed method and state-of-the-art method: canonical correlation analysis (CCA). The higher results are better.

5. RESULTS AND DISCUSSION

In the present study, we set the duration of each trial to seven seconds of blinking for each direction. Then a moving window of two seconds was used with an overlap of 1.98 seconds. Thus, one training sample for a specific control direction (or class) is two seconds, which corresponds to 500 samples for each of the eight channels, as shown in Fig. 2. Each seven seconds resulted in 250 moving windows. For each subject, the total number of the samples for training phase were 15,000 and the same number of samples for testing phase. The accuracy of the classifier was calculated for each subject.

We compared our results to the state-of-the-art method: CCA. As shown in Table 1, the proposed method achieved better performance over the state-of-the-art method. It achieved an average accuracy 92.3% compared to 91.45% for CCA, as shown in Table 1. With the female samples (Subject 4), the CCA method could only achieve 77%. This may have been the results of the high impedance during recording due to long hair, causing a relatively poor quality signal compared to the male data. However, the hybrid architecture has been able to overcome this difficulty. Deep learning is resilient to the noises that are super-imposed on the signal and maintain the same ability to generalize the learned model to un-seen data. For the Subject 2, the classification accuracy is dropped due to the limitation of the RNN layer to capture the higher order harmonics compared to CCA.

6. CONCLUSIONS

In the presented work, we proposed a novel architecture for SSVEP signal classification in the time domain. We utilized in our experiments only eight channels for EEG signal capture procedure. The advantage of the proposed method, is the usage of a raw time-series signal as the input signal. Thus, it can be used for efficient real-time classification of the SSVEP signals. The Separable 2-D convolutional kernel have a superior performance over the regular 2-D convolutional networks that share the kernel weights. The reason for that is the ability to learn better features for each channel independently. Thus, no data normalization is required as a pre-processing step. The preliminary off-line analysis has given very promising results on the classification of SSVEP data using a hybrid architecture incorporating CNN and RNN. One of the limitations is that RNN layer was able to capture only the dominant frequency while the CCA managed to capture both the fundamental frequency and its harmonics. Future work will include the real-time implementation of the network to classify the SSVEP in the real-time to control a robot car and exoskeleton.

7. REFERENCES

[1] Han Yuan and Bin He, "Brain–computer interfaces using sensorimotor rhythms: current state and future perspectives," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 5, pp. 1425–1435, 2014.

[2] Kai Keng Ang and Cuntai Guan, "Eeg-based strategies to detect motor imagery for control and rehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 4, pp. 392–401, 2017.

[3] Christoph Guger, Brendan Z Allison, Bernhard Großwindhager, Robert Prückl, Christoph Hintermüller, Christoph Kapeller, Markus Bruckner, Gunther Krausz, and Günter Edlinger, "How many people could use an ssvep bci?," *Frontiers in neuroscience*, vol. 6, 2012.

[4] Xiaogang Chen, Yijun Wang, Masaki Nakanishi, Xiaorong Gao, Tzyy-Ping Jung, and Shangkai Gao, "High-speed spelling with a noninvasive brain–computer interface," *Proceedings of the national academy of sciences*, vol. 112, no. 44, pp. E6058–E6067, 2015.

[5] Yu Zhang, Guoxu Zhou, Qibin Zhao, Akinari Onishi, Jing Jin, Xingyu Wang, and Andrzej Cichocki, "Multiway canonical correlation analysis for frequency components recognition in ssvep-based bcis," in *Neural Information Processing*. Springer, 2011, pp. 287–295.

[6] Hanh Vu, Bonkon Koo, and Seungjin Choi, "Frequency detection for ssvep-based bci using deep canonical correlation analysis," in *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 001983–001987.

[7] Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu, "Deep canonical correlation analysis," in *International Conference on Machine Learning*, 2013, pp. 1247–1255.

[8] Jie Pan, Xiaorong Gao, Fang Duan, Zheng Yan, and Shangkai Gao, "Enhancing the classification accuracy of steady-state visual evoked potential-based brain–computer interfaces using phase constrained canonical correlation analysis," *Journal of neural engineering*, vol. 8, no. 3, pp. 036027, 2011.

[9] Yu Zhang, Jing Jin, Xiangyun Qing, Bei Wang, and Xingyu Wang, "Lasso based stimulus frequency recognition model for ssvep bcis," *Biomedical Signal Processing and Control*, vol. 7, no. 2, pp. 104–111, 2012.

[10] Yangsong Zhang, Peng Xu, Kaiwen Cheng, and Dezhong Yao, "Multivariate synchronization index for frequency recognition of ssvep-based brain–computer interface," *Journal of neuroscience methods*, vol. 221, pp. 32–40, 2014.

[11] No-Sang Kwak, Klaus-Robert Müller, and Seong-Whan Lee, "A convolutional neural network for steady state visual evoked potential classification under ambulatory environment," *PloS one*, vol. 12, no. 2, pp. e0172578, 2017.

[12] Pouya Bashivan, Irina Rish, Mohammed Yeasin, and Noel Codella, "Learning representations from eeg with deep recurrent-convolutional neural networks," *arXiv preprint arXiv:1511.06448*, 2015.

[13] Jürgen Schmidhuber, Daan Wierstra, and Faustino Gomez, "Evolino: Hybrid neuroevolution/optimal linear search for sequence learning," in *Proceedings of the 19th international joint conference on Artificial intelligence*. Morgan Kaufmann Publishers Inc., 2005, pp. 853–858.

[14] M. Attia, M. Hossny, S. Nahavandi, and A. Yazdabadi, "Skin melanoma segmentation using recurrent and convolutional neural networks," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, April 2017, pp. 292–296.

[15] M. Attia, M. Hossny, S. Nahavandi, and H. Asadi, "Surgical tool segmentation using a hybrid deep cnn-rnn auto encoder-decoder," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, October 2017.